Volume 216, number 2, 229-233 FEB 04763 June 1987

Molecular cloning and sequence analysis of cDNA coding for the precursor of the human cysteine proteinase inhibitor cystatin C

Magnus Abrahamson, Anders Grubb, Isleifur Olafsson and Åke Lundwall

Department of Clinical Chemistry, University of Lund, Malmö General Hospital, S-214 01 Malmö, Sweden

Received 25 March 1987

Recombinant cystatin C producing clones were isolated from a human placenta λgt11 cDNA library. The cDNA insert of one of the clones, containing 777 base pairs, encodes the complete mature cystatin C (120 amino acids) and a hydrophobic leader sequence of 26 amino acids, indicating an extracellular function of the inhibitor. The deduced protein sequence confirms the protein sequence of cystatin C isolated from human urine, but differs in one position from the sequence of the cystatin C fragment deposited as amyloid in hereditary cerebral hemorrhage with amyloidosis.

Cerebral hemorrhage; Cystatin C; Cysteine proteinase; Enzyme inhibitor; cDNA cloning; Nucleotide sequence

1. INTRODUCTION

Cysteine proteinases participate in the intracellular catabolism of proteins and peptides [1], proteolytic processing of prohormones [2], extracellular breakdown of collagen [3] and penetration of normal tissues by malignant cells [4] and probably also by microorganisms [5]. Their physiological activity is regulated by specific protein inhibitors which recently were found to define a new superfamily of proteins called the cystatin superfamily [6]. The high-M_r cystatins, or kininogens, are heavily glycosylated, their mRNAs code for hydrophobic leader sequences [7,8], and one of their key functions is therefore assumed to be extracellular inhibition of cysteine proteinases. Several of the low-M_r cystatins characterized have also been reported to occur in extracellular fluids [9]. However, the low-M_r cystatins are not glycosylated, and since no mRNAs for this group of inhibitors have been described, it has not been ascertained whether any of these are genuine extracellular inhibitors. The present work was undertaken to examine the properties of the mRNA for cystatin C which has the highest concentration of all known low-M_r cystatins in most human extracellular fluids investigated [9] and which plays an important role in the pathophysiology of hereditary cerebral hemorrhage with amyloidosis [10-13]. The amino acid sequence of cystatin C isolated from human urine [13,14] was recently reported to be homologous with c-Ha-*ras* oncogene products [15].

Correspondence address: M. Abrahamson, Department of Clinical Chemistry, University of Lund, Malmö General Hospital, S-21401 Malmö, Sweden

2. MATERIALS AND METHODS

The construction of a human prostate cDNA library has previously been described [16]. λgt11 cDNA libraries made from human placenta and liver mRNA were generous gifts from Drs José Milan (La Jolla Cancer Research Foundation) and Savio Woo (Baylor College of Medicine, Houston, TX), respectively. Affinity purified antibodies against cystatin C were prepared by immuno-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50 © 1987 Federation of European Biochemical Societies

sorption of a polyclonal rabbit antiserum raised against cystatin C isolated from human urine [17]. A mixed oligonucleotide (23-mer) with the sequence AA (T/C)TG(T/C)CC(A/T/C/G)TT-(T/C)CA(T/C) GA(T/C)CA(A/G)CC, based on the amino acid sequence of human cystatin C (amino acids 82–89 [14]), was a kind gift from Dr R. Wydro, Integrated Genetics.

2.1. *Isolation and analysis of cDNA clones*

Recombinant phages were screened at a density of 50 000 plaques per 130 mm petri dish with the affinity-purified antibodies, according to Young and Davis [18,19]. Bound antibodies were detected with alkaline phosphatase-conjugated secondary antibodies (Promega Biotechniques), as described by the manufacturer. λ phages, prepared by the plate lysate method, were isolated by centrifugation in a CsCl gradient [20]. The DNA was extracted by standard methodology [20]. A mixed oligonucleotide probe specific for cystatin C was hybridized to *Eco*RI digests of phage DNA in Southern blot experiments. Hybridizing cDNA inserts were ligated into *Eco*RI-linearized pUC18 plasmid vectors which were subsequently used to transform *E. coli* JM83 cells. Plasmids were prepared by the alkaline lysis method [20]. DNA sequencing of inserts subcloned in M13mp8 was performed using a modified dideoxy chain terminator technique [21]. The insert of the full-length clone C6a was sequenced using the shotgun approach [22]. Nucleotide sequences were aligned and analysed by computer programmes described by Staden [23,24].

2.2. *Southern blotting and plaque hybridizations*

DNA was fractionated by electrophoresis in 0.9% agarose gels and transferred to nitrocellulose filters as described [25]. Hybridizations with the mixed oligonucleotide probe, 5′-end labelled with ^{32}P, were carried out at 42°C for 16 h in a solution containing 6 × SSC (1 × SSC = 150 mM NaCl/15 mM sodium citrate, pH 7.0), 1 × Denhardt's solution (0.02% bovine serum albumin/0.02% polyvinylpyrrolidone/0.02% Ficoll) and 0.05% sodium pyrophosphate. Filters were washed with several changes of 6 × SSC/0.05% sodium pyrophosphate at room temperature, then at 37°C for 1 h and at 50°C for 10 min. Hybridizations with nick-translated cDNA were done as described [20]. A commercial kit (Amersham) was used for the nick translations, according to the manufacturer's instructions.

3. RESULTS

cDNA libraries made from human liver, prostate and placenta were screened for cystatin C encoding clones. Approx. 1.2×10^6 recombinants of the liver library and 3×10^5 recombinants of the prostate library were screened with both antibodies and a mixed oligonucleotide probe. The outcome of this screening was negative. In contrast, screening of 6×10^5 recombinants of the placenta library yielded nine clones reacting with the antibody. Three of these clones, designated C5, C6a and C12, were further processed. Their inserts of about 800, 800 and 700 base pairs, respectively, hybridized specifically in Southern blot experiments to a mixed oligonucleotide constructed from protein sequence data. Sequencing of inserts from clones C6a and C12 subcloned in *Eco*RI-digested M13mp8 revealed that the inserts shared the same 3′ sequence, containing the polyadenylation signal AATAAA. The entire sequence of both strands of the clone C6a cDNA insert was determined by sequencing randomly generated overlapping fragments subcloned in M13mp8. Each nucleotide was determined on average 5.63 times. Furthermore, the data obtained when sequencing the cDNA in-

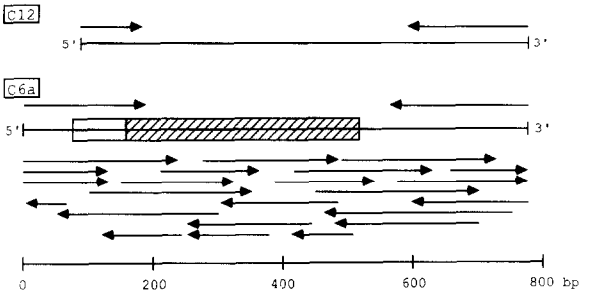

Fig.1. Sequencing strategy for the human cystatin C cDNA inserts. The inserts of clones C12 and C6a were sequenced from the ends by the chain terminator method. A shotgun sequencing analysis technique was used to sequence both DNA strands of the C6a insert. Horizontal arrows indicate the direction and extent of each sequence analysis. The protein coding region of insert C6a is boxed, the hatched part showing the region coding for the mature protein.

```
                                                                    22
                                                  GGGCGCAGCGGGTCCTCTCTAT
                                                                    95
CTAGCTCCAGCCTCTCGCCTGCGCCCCACTCCCCGCGTCCCGCTCCTAGCCGACC ATG GCC GGG CCC CTG CGC
                                                  Met Ala Gly Pro Leu Arg
                                                  -26                -21
                                                                   155
GCC CCG CTG CTC CTG CTG GCC ATC CTG GCC GTG GCC CTG GCC GTG AGC CCC GCG GCC GGC
Ala Pro Leu Leu Leu Leu Ala Ile Leu Ala Val Ala Leu Ala Val Ser Pro Ala Ala Gly
                                                                    -1
                                                                   215
TCC AGT CCC GGC AAG CCG CCG CGC CTG GTG GGA GGC CCC ATG GAC GCC AGC GTG GAG GAG
Ser Ser Pro Gly Lys Pro Pro Arg Leu Val Gly Gly Pro Met Asp Ala Ser Val Glu Glu
 1                                                                  20
                                                                   275
GAG GGT GTG CGG CGT GCA CTG GAC TTT GCC GTC GGC GAG TAC AAC AAA GCC AGC AAC GAC
Glu Gly Val Arg Arg Ala Leu Asp Phe Ala Val Gly Glu Tyr Asn Lys Ala Ser Asn Asp
                                                                   40
                                                                   335
ATG TAC CAC AGC CGC GCG CTG CAG GTG GTG CGC GCC CGC AAG CAG ATC GTA GCT GGG GTG
Met Tyr His Ser Arg Ala Leu Gln Val Val Arg Ala Arg Lys Gln Ile Val Ala Gly Val
                                                                   60
                                                                   395
AAC TAC TTC TTG GAC GTG GAG CTG GGC CGA ACC ACG TGT ACC AAG ACC CAG CCC AAC TTG
Asn Tyr Phe Leu Asp Val Glu Leu Gly Arg Thr Thr Cys Thr Lys Thr Gln Pro Asn Leu
                                                                   80
                                                                   455
GAC AAC TGC CCC TTC CAT GAC CAG CCA CAT CTG AAA AGG AAA GCA TTC TGC TCT TTC CAG
Asp Asn Cys Pro Phe His Asp Gln Pro His Leu Lys Arg Lys Ala Phe Cys Ser Phe Gln
                                                                   100
                                                                   515
ATC TAC GCT GTG CCT TGG CAG GGC ACA ATG ACC TTG TCG AAA TCC ACC TGT CAG GAC GCC
Ile Tyr Ala Val Pro Trp Gln Gly Thr Met Thr Leu Ser Lys Ser Thr Cys Gln Asp Ala
                                                                   120
                                                                   593
TAG GGGTCTGTACCGGGCTGGCCTGTGCCTATCACCTCTTATGCACACCTCCCACCCCCTGTATTCCCACCCCTGGAC

                                                                   672
TGGTGGCCCCTGCCTTGGGGAAGGTCTCCCCATGTGCCTGCACCAGGAGACAGACAGAGAAGGCAGCAGGCGGCCTTTG

                                                                   751
TTGCTCAGCAAGGGGCTCTGCCCTCCCTCCTTCCTTCTTGCTTCTCATAGCCCCGGTGTGCGGTGCATACACCCCCACC
                                777
TCCTGCAATAAAATAGTAGCATCCCC
```

Fig.2. Nucleotide and deduced amino acid sequence of clone C6a cDNA encoding human precystatin C. Numbering of the nucleotide sequence begins at first nucleotide and proceeds in the 5′ to the 3′ direction. Amino acid numbering starts with residue 1 of the mature protein. The amino acid substituted in the cystatin C fragment deposited as amyloid in patients with hereditary cerebral hermorrhage with amyloidosis is boxed. The Kozak initiation consensus and the polyadenylation signal are underlined.

serts of clones C6a and C12 from the ends revealed that the sequence of the C12 insert starts at position 91 of that of C6a (fig.1).

The C6a insert contains 777 base pairs, including 77 base pairs of 5'non-coding sequence and 262 base pairs of 3'non-coding sequence (fig.2). The polyadenylation signal AATAAA at positions 756–761 is followed by 15 nucleotides. A probable initiator methionine, located six nucleotides downstream of a stop codon and flanked by sequences in agreement with the Kozak initiation consensus [26], is coded for by nucleotides 78–80. An open reading frame extending from the postulated initiator methionine to a stop codon, TAG, at positions 516–518, encodes the 120 amino acids of cystatin C isolated from urine [13,14] and a putative hydrophobic signal sequence of 26 amino acid residues.

The cDNA insert of clone C6a was labelled by nick translation and used to rescreen 120000 recombinants of the placenta library. Seven of these hybridized to the probe, indicating a frequency of cystatin C mRNA in placenta of less than 0.01%.

4. DISCUSSION

Cystatin C has been demonstrated to occur intracellularly in many neuroendocrine cell types by immunohistochemical procedures [13], but no reports concerning secretion of cystatin C into extracellular fluids have been published. However, seminal plasma has been determined to have the highest concentration of cystatin C of all biological fluids investigated so far [9] and recent studies in our laboratory indicate that fibroblasts and glial cells secrete cystatin C into tissue cultivation fluids. We therefore investigated human prostate and placenta cDNA libraries, and in addition a human liver cDNA library, in an effort to find clones carrying inserts complementary to cystatin C mRNA. Such clones could only be identified in the placenta library, however.

The amino acid sequence deduced from the nucleotide sequence of the cloned cystatin C cDNA specifies a precystatin C polypeptide chain containing not only the amino acid sequence of mature cystatin C [13,14] but also a hydrophobic leader sequence of 26 amino acids. It is therefore probable that cystatin C has a physiological function as a cysteine proteinase inhibitor extracellularly. Indeed, a recent investigation of the concentrations of all major cystatins, including the kininogens and cystatin C, in several human extracellular body fluids combined with enzyme kinetic measurements, indicated that cystatin C could be a physiologically important extracellular cysteine proteinase inhibitor in several of the investigated fluids [9].

Cystatin C plays an important role in the pathophysiology of hereditary cerebral hermorrhage with amyloidosis in which it is deposited as amyloid in the walls of the rupturing cerebral arteries [9,10]. The metabolism of cystatin C in this disease, which affects young adults, is altered [12] and the sequence of the deposited cystatin C differs in one position from that originally described for cystatin C isolated from persons without the disease [14,27]. It is interesting that the cDNA presently described specifies a mature cystatin C with an amino acid sequence identical to that originally described for 'normal' cystatin C and that the difference (Gln instead of Leu at position 68) can be explained by a single base substitution giving a CAG instead of a CTG sequence at positions 357–359. The cloned cDNA for normal cystatin C can hopefully be used as a probe in elucidating the structure of the gene for the normal inhibitor as well as the genetic background to hereditary cerebral hemorrhage with amyloidosis.

ACKNOWLEDGEMENTS

This work was supported by A. Österlunds Stiftelse, Direktör A. Påhlssons Stiftelse, G. och J. Kocks Stiftelse, the Medical Faculty, University of Lund, and the Swedish Medical Research Council (project no. 05196).

REFERENCES

[1] Barrett, A.J. and Kirschke, H. (1981) Methods Enzymol. 80, 535–561.
[2] Marks, N., Berg, M.J. and Benuck, M. (1986) Arch. Biochem. Biophys. 249, 489–499.
[3] Delaissé, J.-M., Eeckhout, Y. and Vaes, G. (1984) Biochem. Biophys. Res. Commun. 125, 441–447.
[4] Poole, A.R. (1973) in: Lysosomes in Biology and Pathology (Dingle, J.T. ed.) vol. 3, pp. 303–307, North-Holland, Amsterdam.

[5] Barrett, A.J., Davies, M.E. and Grubb, A. (1984) Biochem. Biophys. Res. Commun. 120, 631–636.
[6] Barrett, A.J., Fritz, H., Grubb, A., Isemura, S., Järvinen, M., Katunuma, N., Machleidt, W., Müller-Esterl, W., Sasaki, M. and Turk, V. (1986) Biochem. J. 236, 213.
[7] Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H. and Sasaki, M. (1984) Biochemistry 23, 5691–5697.
[8] Takagaki, Y., Kitamura, N. and Nakanishi, S. (1985) J. Biol. Chem. 260, 8601–8609.
[9] Abrahamson, M., Barrett, A.J., Salvesen, G. and Grubb, A. (1986) J. Biol. Chem. 261, 11282–11289.
[10] Gudmundsson, G., Hallgrimsson, J., Jonasson, T.A. and Bjarnason, O. (1972) Brain 95, 387–404.
[11] Cohen, D.H., Feiner, H., Jensson, O. and Frangione, B. (1983) J. Exp. Med. 158, 623–628.
[12] Grubb, A., Jensson, O., Gudmundsson, G., Arnason, A., Löfberg, H. and Malm, J. (1984) N. Engl. J. Med. 311, 1547–1549.
[13] Grubb, A. and Löfberg, H. (1985) Scand. J. Clin. Lab. Invest. 45, Suppl. 177, 7–13.
[14] Grubb, A. and Löfberg, H. (1982) Proc. Natl. Acad. Sci. USA 79, 3024–3027.
[15] Hiwasa, T., Yokoyama, S., Ha, J.-M., Noguchi, S. and Sakiyama, S. (1987) FEBS Lett. 211, 23–26.
[16] Lundwall, Å. and Lilja, H. (1987) FEBS Lett. 214, 317–322.
[17] Löfberg, H., Grubb, A.O. and Brun, A. (1981) Biomed. Res. 2, 298–306.
[18] Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194–1198.
[19] Young, R.A. and Davis, R.W. (1983) Science 222, 778–782.
[20] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
[21] Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.
[22] Bankier, A.T. and Barrell, B.G. (1983) in: Techniques in Nucleic Acid Biochemistry (Flavell, R.A. ed.) vol. 85, pp. 1–73, Elsevier, Limerick, Ireland.
[23] Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961.
[24] Staden, R. (1982) Nucleic Acids Res. 10, 4731–4751.
[25] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[26] Kozak, M. (1981) Nucleic Acids Res. 9, 5233–5252.
[27] Ghiso, J., Jensson, O. and Frangione, B. (1986) Proc. Natl. Acad. Sci. USA 83, 2974–2978.